

2015年度 修士論文

金融市場に対する外生的ショックの  
実証研究

Empirical study of exogenous shocks in  
financial markets

中央大学大学院 理工学研究科  
博士課程前期課程 物理学専攻  
田平 好文

## 概要

本論文では、株式市場と国債に対する外生的ショックとして、それぞれニュースと日本銀行の国債買入について調査する。

まず、2007年12月から2012年4月までの THOMSON REUTERS によって配信された全ての英語ニュース記事(915万記事)を用いて、ある株式市場に関連するニュース記事数の変動と将来のその市場の株式指数の変動に相関があることを明らかにする。ある週にニューヨーク証券取引所(NYSE)と全米証券業協会自動株相場(NASDAQ)に上場する銘柄に関連するニュース記事数が増加/減少すると、その翌週にアメリカの株式市場の動向を示す指数である Standard & Poors 500 index (S&P500 index) は下降/上昇する傾向にある。しかし、NYSE や NASDAQ と関連の低い他の市場の株式指数の変動はランダムに近い。特に、企業の業績と密接に関係する Business sector のニュース記事数の変動には、市場は顕著に反応する。これらの傾向は各年で観測される。

次に、日本銀行の国債買入オペレーションには、日本国債の価格を上昇させる効果があり、それによりドル円レートも変動することを明らかにする。日米間の国債の金利差とドル円レートの対数変動の相関は強いが、日本国債とドル円レートの相関は弱い。しかし、日本銀行による国債買入のアナウンスが行われると、市場における日本国債の売り注文が減り、日本国債の価格は上昇(金利下降)する。また、ドル円レートも円高方向に変動し、日本国債とドル円レートの相関が強くなる。

# 目次

1	はじめに	3
2	金融市場の方向性とニュース報道の量的パターン	5
2.1	ニュースデータ	5
2.2	ニュースを利用した売買戦略	6
2.3	ランダムな売買戦略	7
2.4	時系列解析	8
2.5	ニュース記事数の変動と株価指数の変動の関係性	10
2.6	累積収益率のニューストピックへの依存性	12
2.7	累積収益率の時期依存性	16
2.8	日本国債への応用	17
2.9	まとめ	17
3	日本国債・米国債・ドル円レートの3体相関	19
3.1	日本国債の先物と現物の連動性	19
3.2	日本国債先物の統計性	21
3.3	日本国債・米国債とドル円レートの関係	23
3.4	日本国債とドル円レートに対する日本銀行の国債買入の影響	25
3.5	まとめ	29
付録 A	株式銘柄に関するニュースの分類	30

# 1 はじめに

金融市場は非平衡開放系とみなすことができ、経済や経済物理等の分野で研究されている。金融市場ではしばしば正規分布から外れた激しい価格変動が起き、市場経済を混乱させている。このような価格変動は、内生的ショックと外生的ショックによって起きている [1][2]。過去の取引価格を記録したティックデータの分析によって、主要な内生的ショックが乗算過程に従っていることが明らかにされた。乗算過程は価格変動分布のファットテイルを生み出す [3]。

外生的ショックについても研究されており、その代表としてニュース報道がある [4][5][6][7][8][9]。ニュース本文に対する自然言語処理により、ニュースのセンチメントやトピックを抽出し、それらと価格変動に相関があることが示された [10][11][12]。さらに、ニュースの内容そのものではなく、ニュースに対する人々の反応と価格変動の関係についても調査されている。

行動経済学におけるプロスペクト理論では、人々は株の売買により利益を得ることよりも損失を回避することを優先的に考えると仮定する [13]。この仮定に従えば、人々は株価に下落のおそれがある時、その株に関連する話題に関して Web 上で必死に情報収集を行い、下落による損失を防ごうとする。つまり、株価が下落する前にその株に関連する話題の Web 上での検索数・閲覧数の増加が想定される。T. Preis らは、Dow Jones Industrial Average (DJIA) の構成銘柄である 30 社についての Web 上での検索数・閲覧数の変動とその後の DJIA の変動に関連があり、株価の下落の前に上記の話題に関する検索数・閲覧数が増加することを示した [14][15][16]。

本論文の 2 章では、ある株式市場に上場する企業についてのニュースの記事数の変動とその後のその市場の株式指数の変動に関係があることを示す。様々な報道機関は人々が注目するであろう話題を中心に報道するため、Web 上での検索数・閲覧数と同様に株式指数の下落の前に特定の話題に関するニュースの記事数が増加すると考えられる。この仮定をビジネスニュースの記事数と株式指数の時系列を用いて証明する [17]。

ニュース以外の外生的ショックとして、主要銀行による市場介入等のマーケットインパクトがある [18]。経済物理学では、株式市場や為替市場を対象とした上記の研究は多くされているが、国債を対象とした研究はほとんどされていない。3 章では、日本銀行の国債買入による日本国債の価格変動について述べる。まず国債の価格変動の統計性が、株式市場や為替市場でみられる統計性と一致することを確認する。次に、国債の価格 (金利) 変動が与える影響として日本国債、米国債、ドル円レート of 相関に注目する。経済学における金利平価説では、2 国間で金利が高い方の国の通貨が将来的に減価し、金利が低い方の国の通貨が増価することが言われている。たしかに、日本国債と米国債の金利差とドル円レートの変動に強

い相関があるが、日本国債がドル円レートに与える影響は弱く、米国債が与える影響の方が強い。ここで、日本銀行の国債買入が行われた日に注目すると、日本国債に価格上昇(金利下降)のトレンド、ドル円レートに円高方向へのトレンドが発生し、相関が強くなることを示す。

## 2 金融市場の方向性とニュース報道の量的パターン

2章は次のように構成される。2.1節では、本論文で扱うニュースデータについて述べる。2.2節では、ニュースの記事数の変動と株価指数の変動の関係を調べる手法について述べる。2.3節では、ランダムな売買戦略によって得られる累積収益率の期待値と標準偏差を解析的に計算する。2.4節では、ニューヨーク証券取引所 (NYSE) に上場する銘柄の1つである Schlumberger に関するニュースの記事数の変動とアメリカの株式市場の動向を示す指数である Standard & Poors 500 index (S&P500 index) の変動に関する関係があることを示す。2.5節では、NYSE と全米証券業協会自動株相場 (NASDAQ) の上場銘柄に関するニュースの記事数の変動と S&P500 index の変動に関する関係があることを示す。2.6節と 2.7節では、この関係のニューストピックおよび、時期に対する依存性を調査する。

### 2.1 ニュースデータ

本論文では、世界最大の金融情報サービス THOMSON REUTERS 社が機関投資家向けに 2007 年 12 月から 2012 年 4 月の間に配信した英語のニュース 915 万記事を用いる [19][20]。ニュース記事は ALERT と HEADLINE と STORY\_TAKE\_OVERWRITE の 3 種類に分類される。ALERT は緊急性の高いニュースで、記事にはタイトルのみが含まれる。HEADLINE はその他のニュースで、同じく記事にはタイトルのみが含まれる。STORY\_TAKE\_OVERWRITE は ALERT と HEADLINE が配信された後に配信されるそれらの記事の本文である。実質的なニュース数を数えるために ALERT と HEADLINE に分類される 915 万記事のみを使う。

それぞれのニュース記事にはトピックコードと呼ばれる記者によってつけられたキーワードが付随する。各トピックコードは “Arts / Culture / Entertainment”, “Asset Class / Property”, “Business sector”, “Commodity”, “Crime”, “Currency”, “Disaster / Accident”, “Event Type”, “Genre”, “Geography”, “Health / Medicine”, “Indicator Type”, “Intellectual Property”, “Language”, “Legacy News Topic”, “News Flag / Status”, “Organization”, “Physical Asset Type”, “Religion”, “Sport”, “Sport combined with Geography”, “Sporting Competition” の 22 分類に属する。また、それぞれのニュースには銘柄コードと呼ばれる株式銘柄を示すコードもある。

本論文では、2007 年 12 月から 2012 年 4 月の 230 週の間にはニュース記事が 130 週より多く存在した株式銘柄・トピックのみを分析に利用した。この条件に該当した NYSE と NASDAQ の銘柄数とトピック数はそれぞれ 500 銘柄と 676 個である。他の報道機関のニュースにおける株式銘柄・トピックの分類に関しては付録 A を参照。

## 2.2 ニュースを利用した売買戦略

T. Preis らの Google の検索数・Wikipedia の閲覧数にもとづいた売買戦略 [14][15][16] を、ニュース記事数に応用することによって、我々はある特定の株式市場に関係するニュース記事数の変動とその後の株価指数の変動との関係を明らかにする。

ある週  $t-1$  の上場企業  $i$  に関するニュースの記事数を  $n_i(t-1)$  とする。ここで、週の終わりは日曜日とする。次に、過去  $\Delta t$  週間の記事数の平均

$$N_i(t-2, \Delta t) = \frac{1}{\Delta t} \sum_{\tau=2}^{\Delta t+1} n_i(t-\tau) \quad (1)$$

を計算する。週  $t-1$  の記事数が過去より増加した  $n_i(t-1) > N_i(t-2, \Delta t)$  とき、 $t$  週のはじめにその企業が上場している株式市場における 1 単位の株式指数を  $p(t)$  で売り、 $t+1$  週のはじめに  $p(t+1)$  で買い戻す。このとき  $x(t) = +1$  とすると、対数収益率は

$$x(t)(\ln p(t) - \ln p(t+1)) = \ln p(t) - \ln p(t+1) \quad (2)$$

で与えられる\*1。逆に、週  $t-1$  の記事数が過去より減少した  $n_i(t-1) \leq N_i(t-2, \Delta t)$  とき、 $t$  週のはじめに 1 単位のその株式指数を  $p(t)$  で買い、 $t+1$  週のはじめに  $p(t+1)$  で同じだけ売る。このとき  $x(t) = -1$  とすると、対数収益率は

$$x(t)(\ln p(t) - \ln p(t+1)) = \ln p(t+1) - \ln p(t) \quad (3)$$

で与えられる。この仮想的な売買戦略を特定の期間  $T$  で行い、累積収益率

$$R_i(T) = \sum_{t=1}^T x(t)(\ln p(t) - \ln p(t+1)) \quad (4)$$

を計算する。この累積収益率  $R_i$  をランダムに売買を選択する戦略で得られる累積収益率の標準偏差で規格化する。

ある株式市場に上場する全ての企業について規格化された累積収益率  $R_i$  を算出する。 $R_i$  の分布が、ニュースの記事数の変動に依存せずにランダムに売買を選択する戦略で得られる累積収益率の分布と異なることを示すことによって、ニュース記事数の変動がその後の株式指数の変動と有意に関係していることを明らかにすることができる。

---

\*1 株の売買等による、投資額に対する収益の割合は一般的に、収益率  $\frac{p(t)-p(t+1)}{p(t+1)}$  か、対数収益率  $\ln \frac{p(t)}{p(t+1)} = \ln(1 + \text{収益率})$  で計算される。本論文では、Wikipedia の閲覧数にもとづいた売買戦略 [15] との比較のため、[15] と同様に対数収益率を計算する。

## 2.3 ランダムな売買戦略

ランダムな売買戦略の累積収益率の期待値と標準偏差を解析的に計算する．ランダムな売買戦略の  $T$  週間の累積収益率  $R(T)$  は以下の式で記述できる．

$$R(T) = \sum_{t=1}^T x(t)(\ln X(t) - \ln X(t+1)) \quad (5)$$

ここで、 $X(t)$  は株式指数、また  $x(t)$  は  $+1$  の値と  $-1$  の値を等確率でとる乱数とする．まず、期待値を求める．

$$E[R(T)] = E\left[\sum_{t=1}^T x(t)(\ln X(t) - \ln X(t+1))\right] \quad (6)$$

$X(t)$  と  $x(t)$  は独立なので、(6) 式は

$$E[R(T)] = \sum_{t=1}^T E[x(t)]E[\ln X(t) - \ln X(t+1)]$$

となり、 $E[x(t)] = 0$  より、

$$E[R(T)] = 0 \quad (7)$$

となる．次に、標準偏差を求める．

$$\begin{aligned} E[R(T)^2] &= E\left[\sum_{t=1}^T x(t)(\ln X(t) - \ln X(t+1)) \times \sum_{s=1}^T x(s)(\ln X(s) - \ln X(s+1))\right] \\ &= \sum_{t=1}^T \sum_{s=1}^T E[x(t)x(s)]E[\ln X(t) \ln X(s)] \\ &\quad - \sum_{t=1}^T \sum_{s=1}^T E[x(t)x(s)]E[\ln X(t) \ln X(s+1)] \\ &\quad - \sum_{t=1}^T \sum_{s=1}^T E[x(t)x(s)]E[\ln X(t+1) \ln X(s)] \\ &\quad + \sum_{t=1}^T \sum_{s=1}^T E[x(t)x(s)]E[\ln X(t+1) \ln X(s+1)] \end{aligned} \quad (8)$$

ここで、 $E[x(t)x(s)] = \begin{cases} 1 & (t = s) \\ 0 & (t \neq s) \end{cases}$  なので、(8) 式は

$$E[R(T)^2] = \sum_{t=1}^T E[\ln X(t)^2] - 2 \sum_{t=1}^T E[\ln X(t) \ln X(t+1)] + \sum_{t=1}^T E[\ln X(t+1)^2] \quad (9)$$

となる。ここで、 $X(t)$  が幾何ブラウン運動過程に従っていると仮定すると、 $X(t)$  は以下の式で記述できる\*<sup>2</sup>。

$$\ln X(t) = \ln X(1) + (\mu - \sigma^2/2)t + \sigma B(t) \quad (10)$$

ここで、 $\mu$  はドリフト、 $\sigma$  はボラティリティと呼ばれるパラメータ、 $B(t)$  は標準ブラウン運動過程である。 $\alpha = \mu - \sigma^2/2$  とおく。(10) 式より

$$E[\ln X(t)^2] = \alpha^2 t^2 + (2\alpha \ln X(1) + \sigma^2)t + \ln X(1)^2 \quad (11)$$

$$\begin{aligned} E[\ln X(t) \ln X(t+1)] &= \alpha^2 t^2 + (2\alpha \ln X(1) + \sigma^2 + \alpha^2)t + \ln X(1)^2 + \alpha \ln X(1) \\ &= E[\ln X(t)^2] + \alpha^2 t + \alpha \ln X(1) \end{aligned} \quad (12)$$

となる。(11) 式と (12) 式を (9) 式に代入し、計算すると以下の式になる。

$$E[R(T)^2] = (\alpha^2 + \sigma^2)T \quad (13)$$

(7) 式と (13) 式より、累積収益率の標準偏差は

$$\sqrt{E[R(T)^2] - E[R(T)]^2} = \sqrt{(\alpha^2 + \sigma^2)T} \quad (14)$$

となる。(7) 式と (14) 式より、ランダムに売買を選択する戦略の累積収益率の期待値は 0 であり、標準偏差は期間  $T$  の 0.5 乗に比例することが示せた。

## 2.4 時系列解析

2.3 節で得られたランダムな売買戦略の累積収益率の理論式と、S&P500 index のデータにもとづいて仮想的にランダムな売買を行った場合の累積収益率を比較する。図 1 は S&P500 index の時系列である。図 2 の黒の実線は、(7) 式と (14) 式をプロットしたものである\*<sup>3</sup>。黒の破線は S&P500 index のデータにもとづいて仮想的にランダムな売買を行った場合の累積収益率の平均値と標準偏差をプロットしたものである。2008 年 9 月から 2009 年 3 月の期間など、S&P500 index の変動が激しい期間を除けば、(14) 式によって得られる累積収益率の標準偏差と S&P500 index をランダムに売買することで得られる累積収益率の標準偏差はほぼ一致している。

ニュースを利用した売買戦略とランダムな売買戦略の比較の 1 つの例として、NYSE に上場する銘柄の 1 つである Schlumberger (SLB) に関する過去  $\Delta t = 7$  週間のニュース記事数の変動にもとづいた売買戦略と、ランダムな売買戦略の累積収益率を比較する。どちらも過去の S&P500 index のデータにもとづき仮想的に売買する。図 2 の青線は

\*<sup>2</sup> 幾何ブラウン運動過程は金融市場の価格変動に対するモデルとしてよく使われる [21]。

\*<sup>3</sup> 最小二乗法によるフィッティングにより  $\sqrt{\alpha^2 + \sigma^2} = 0.03914$  としてプロットした。

Schlumberger に関するニュースを利用した売買戦略の累積収益率の時系列で、累積収益率  $R_{SLB} = 1.77$  となっている。一方、ランダムな売買戦略の累積収益率の平均値と標準偏差は、 $\mu \pm \sigma = 4.04 \times 10^{-4} \pm 0.54$  となっており、ニュースを利用した売買戦略の累積収益率の方が高い。S&P500 index に大きな下落があった 2008 年 9 月から 2009 年 3 月の期間と 2011 年 8 月から 2011 年 10 月の期間で、特に累積収益率が高くなっている。

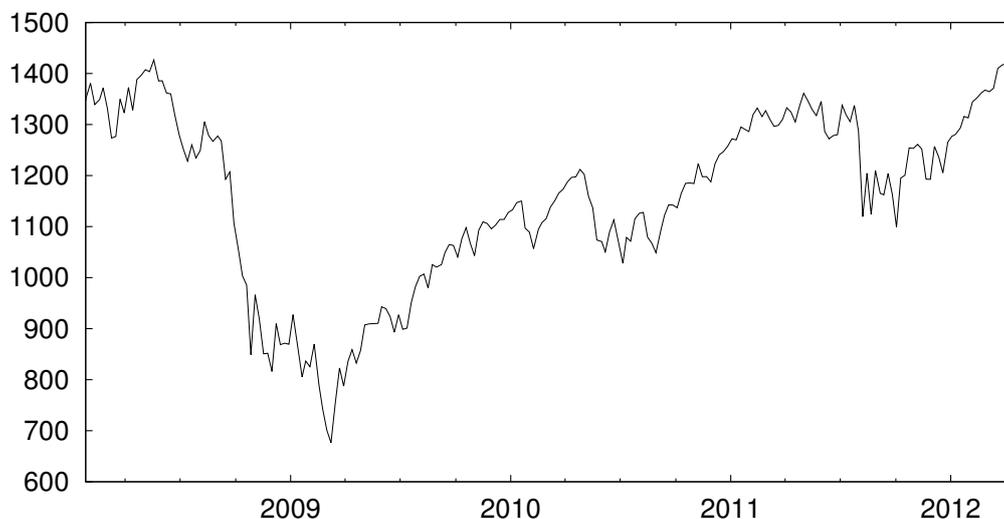


図 1 S&P500 index の週次時系列

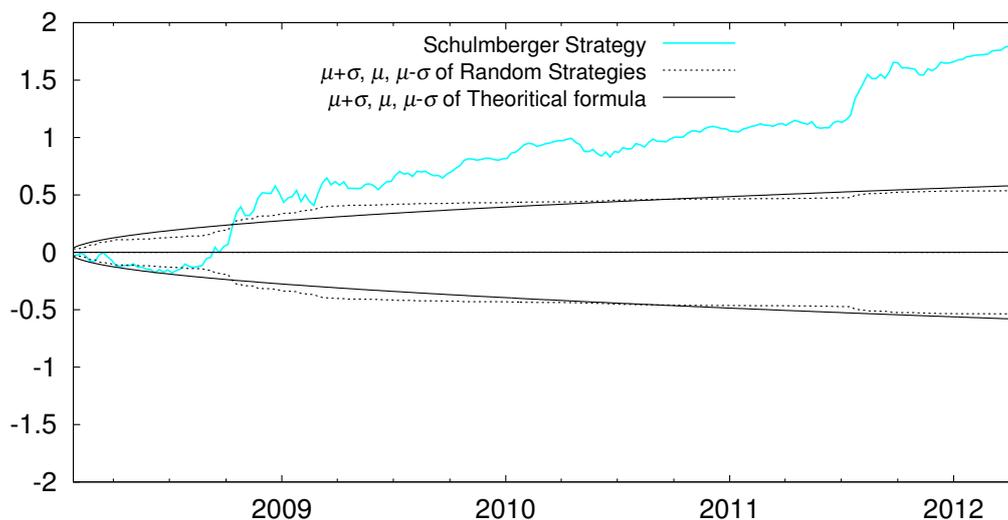


図 2 累積収益率の時系列 青線は Schlumberger に関するニュース記事数の変動にもとづいて S&P500 index を仮想的に売買したときの累積収益率を示す。売買戦略のパラメータ  $\Delta t$  に 7 週をセットした。黒の破線はランダムな売買戦略の累積収益率の平均値と標準偏差を示す。黒の実線は (7) 式と (14) 式をプロットしたものである。

## 2.5 ニュース記事数の変動と株価指数の変動の関係性

2.2 節で述べた売買戦略を用いることで、ある特定の株式市場に関係するニュース記事数の変動とその後の株価指数の変動に関係があることを示す。NYSE, NASDAQ, 東京証券取引所, 香港証券取引所, 上海証券取引所, および韓国証券取引所に上場する銘柄に関するニュース記事を利用して、過去の S&P500 index と日経 225 指数を 2.2 節で述べた売買戦略で仮想的に売買を行う。

図 3(a) は NYSE と NASDAQ に上場する銘柄に関する過去  $\Delta t = 7$  週間のニュース記事数の変動にもとづいて、S&P500 index に対して売買戦略を実行したときの累積収益率  $R_i$  の確率密度分布である。ランダムな売買戦略での平均累積収益率  $1.3 \times 10^{-3}$  に対して、これらの市場の上場企業で平均したニュース記事数の変動にもとづいた売買戦略の累積収益率  $\langle R \rangle$  は高く 0.40 である。2 つの確率密度分布の位置および形状が異なるかどうかを調べる Kolmogorov-Smirnov 検定を行う。Kolmogorov-Smirnov 検定の有意確率  $p$  の値は以下のよう求められる。まず、両側検定統計量

$$D = \sup_x |F_{n_1}(x) - G_{n_2}(x)|$$

を計算する。ここで、 $F_{n_1}$  と  $G_{n_2}$  は、それぞれサンプル数  $n_1$  と  $n_2$  の累積分布である。次に、

$$\chi_0^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2}$$

を計算する。サンプル数  $n_1, n_2$  が十分に大きいとき、 $\chi_0^2$  は近似的に自由度 2 の  $\chi^2$  分布に従う。よって、有意確率  $p$  は

$$p = 2 \times P\{\chi^2 \geq \chi_0^2\}$$

となる。ここで、 $P\{\chi^2 \geq \chi_0^2\}$  は、自由度 2 の  $\chi^2$  分布の右側確率である。図 3(a) における Kolmogorov-Smirnov 検定の  $p$  値は  $1.58 \times 10^{-11}$  で、帰無仮説「ニュースを利用した売買戦略により得られる分布とランダムな売買戦略により得られる分布は同じである」は棄却される。一方で、図 3(b) は NYSE と NASDAQ に上場する銘柄に関する過去  $\Delta t = 7$  週間のニュース記事数の変動にもとづいて、日経 225 指数に対して売買戦略を実行したときの累積収益率  $R_i$  の確率密度分布である。平均累積収益率  $\langle R \rangle$  は低く 0.17 である。そして、Kolmogorov-Smirnov 検定の  $p$  値は 0.041 であり帰無仮説は有意水準 5% で棄却されるが、S&P500 index での  $p$  値  $1.58 \times 10^{-11}$  よりは明らかに高く、よりランダムな売買戦略の分布に近いことが分かる。

次に、異なる株式市場に関する過去  $\Delta t = 7$  週間のニュース記事数の変動にもとづいた売買戦略を S&P500 index に対して実行する。東京証券取引所, 香港証券取引所, 上海証券取引所, および韓国証券取引所に上場する銘柄に関するニュースを利用した売買戦略では平均

累積収益率  $\langle R \rangle$  は 0.13 であり, NYSE と NASDAQ に関するニュースを利用した売買戦略に比べて顕著に低い. また, Kolmogorov-Smirnov 検定の  $p$  値は 0.15 で帰無仮説は有意水準 5% で棄却されない. これらの結果は, 特定の株式指数の将来の変動は, 関係のより深い市場のニュース記事数の変動と相関があることを意味している.

平均累積収益率  $\langle R \rangle$  の  $\Delta t$  への依存性も調査した. 図 4 は過去  $\Delta t$  週間の NYSE・NASDAQ に関するニュース記事数の変動にもとづいて, S&P500 index と日経 225 指数を仮想的に売買した場合の平均累積収益率  $\langle R \rangle$  である. ランダムな売買戦略に比べてニュースを利用した売買戦略の平均累積収益率  $\langle R \rangle$  は  $\Delta t$  に依存せず常に高い. また, 全ての  $\Delta t$  について  $\langle R \rangle$  は S&P500 index の方が日経 225 指数よりも高い.

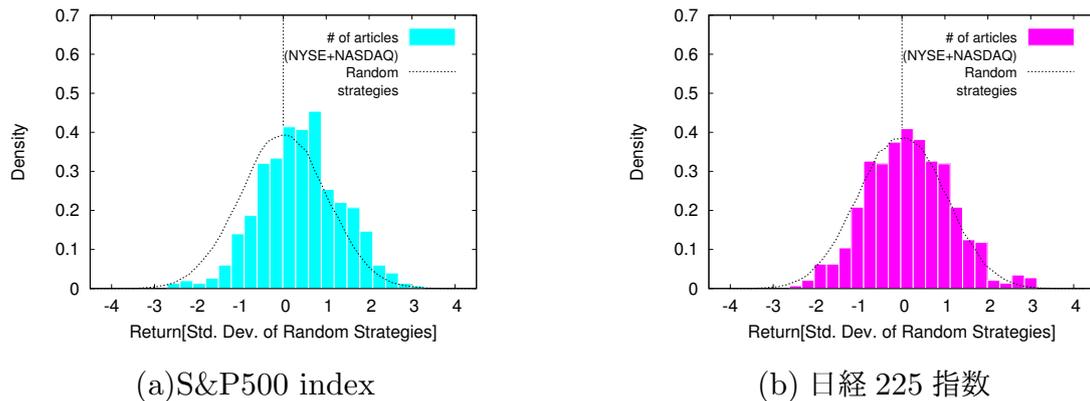


図 3 NYSE と NASDAQ の上場企業に関するニュース記事数の変動にもとづいて (a) S&P500 index と (b) 日経 225 指数を仮想的に売買したときの累積収益率  $R_i$  の確率密度分布 破線の分布はランダムな売買による累積収益率  $R_i$  の分布を示す. 売買戦略のパラメータ  $\Delta t$  に 7 週をセットした. S&P500 index と日経 225 指数を仮想的に売買したときの平均累積収益率  $\langle R \rangle$  は, それぞれ 0.40 と 0.17 である. 破線の分布を帰無仮説とした Kolmogorov-Smirnov 検定の  $p$  値は, それぞれ  $1.58 \times 10^{-11}$  と 0.041 である.

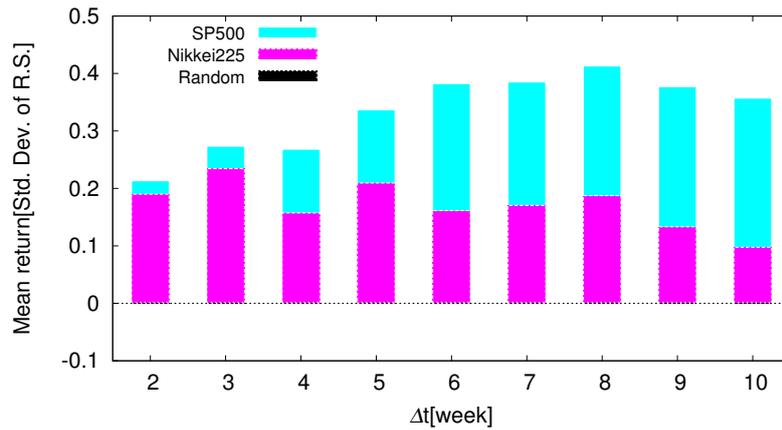


図4 NYSE と NASDAQ の上場銘柄に関する過去  $\Delta t$  週間のニュースを利用した売買戦略による平均累積収益率  $\langle R \rangle$  青と赤は、それぞれ、S&P500 index と日経 225 指数を仮想的に売買したときの平均累積収益率  $\langle R \rangle$  を表す。ほとんど  $\langle R \rangle = 0$  を示す黒は、ランダムに売買したときの平均累積収益率である。

## 2.6 累積収益率のニューストップックへの依存性

各トピックの過去  $\Delta t = 7$  週間のニュース記事数の変動にもとづいて売買戦略を実行することにより、どのようなトピックが将来の株式指数の変動と関係があるかを明らかにする。2.2 節で述べたトピックの分類ごとに、ニュース記事数と S&P500 index を用いて平均累積収益率  $\langle R \rangle$  を算出する\*4。

表 1 は各トピック分類に含まれるトピック数とそれらの平均累積収益率  $\langle R \rangle$  を表す。平均累積収益率  $\langle R \rangle$  は Business sector が一番高く 1.10 で、Sport & Sporting competition が一番低く 0.30 である。図 5 はそれぞれ、Business sector と Sport & Sporting competition の累積収益率の確率密度分布である。Business sector の累積収益率の分布は、ランダムな売買の累積収益率の分布と有意に異なり、Kolmogorov-Smirnov 検定の  $p$  値は  $1.57 \times 10^{-33}$  を示す。一方、Sport & Sporting competition の累積収益率の分布は Kolmogorov-Smirnov 検定の  $p$  値が 0.25 であるから、ランダムな売買の累積収益率の分布に近い。企業の業績と密接に関係する Business sector のニュース記事数の変動には、市場は顕著に反応する。図 6 は Business sector と Sport & Sporting competition に分類されるトピックの累積収益率のランキングである。この図からも、Business sector の累積収益率が顕著に高いことがわ

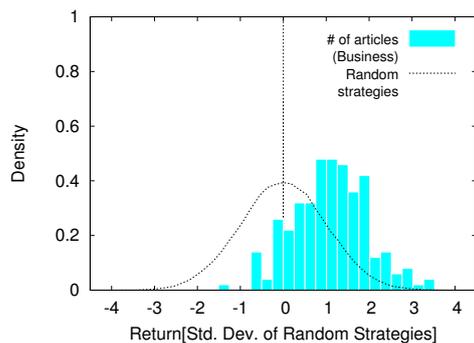
\*4 ただし、トピック数が 10 未満である下記の分類は解析に用いない。（“Genre”, “News Flag / Status”, “Legacy News Topic”, “Indicator Type”, “Arts / Culture / Entertainment”, “Language”, “Health / Medicine”, “Disaster / Accident”, “Physical Asset Type”, “Crime”, “Religion”, “Currency”, “Intellectual Property”, “Sport combined with Geography”）

かる。Business sector でも、表 2 で示される 10 トピックは特に累積収益率が高い。

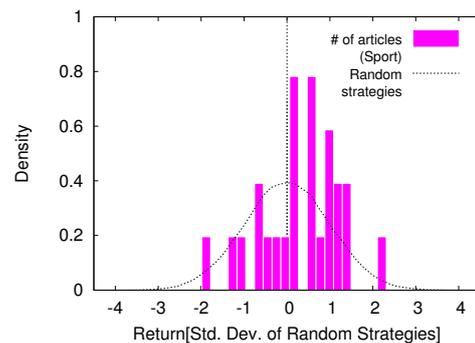
平均累積収益率  $\langle R \rangle$  の  $\Delta t$  への依存性も調査した。図 7 は過去  $\Delta t$  週間の Business sector と Sport & Sporting competition のニュース記事数の変動にもとづいて、S&P500 index を売買した場合の平均累積収益率  $\langle R \rangle$  である。ランダムな売買戦略に比べてニュースを利用した売買戦略の平均累積収益率  $\langle R \rangle$  は  $\Delta t$  に依存せず常に高い。また、全ての  $\Delta t$  について  $\langle R \rangle$  は Business sector の方が Sport & Sporting competition よりも高い。

分類	トピック数	平均累積収益率
Business Sector	199	1.10
Event Type	24	1.08
Commodity	53	0.93
Organization	16	0.46
Asset Class / Property	85	0.37
Geography	207	0.33
Sport & Sporting competition	25	0.30

表 1 トピックの分類におけるトピック数と平均累積収益率 売買戦略のパラメータ  $\Delta t$  に 7 週をセットした。



(a)Business sector



(b)Sport & Sporting competition

図 5 (a)Business sector のニュース記事を利用して S&P500 index を仮想的に売買したときの累積収益率  $R_i$  の確率密度分布 (b)Sport & Sporting competition のニュース記事を利用して S&P500 index を仮想的に売買したときの累積収益率  $R$  の確率密度分布 破線の分布は、ランダムな売買による累積収益率  $R_i$  の分布を示す。売買戦略のパラメータ  $\Delta t$  に 7 週をセットした。Business sector と Sport & Sporting competition のニュース記事を利用して S&P500 index を売買したときの平均累積収益率  $\langle R \rangle$  は、それぞれ 1.10 と 0.30 である。破線の分布を帰無仮説とした Kolmogorov-Smirnov 検定の  $p$  値は、それぞれ、 $1.57 \times 10^{-33}$  と 0.25 である。

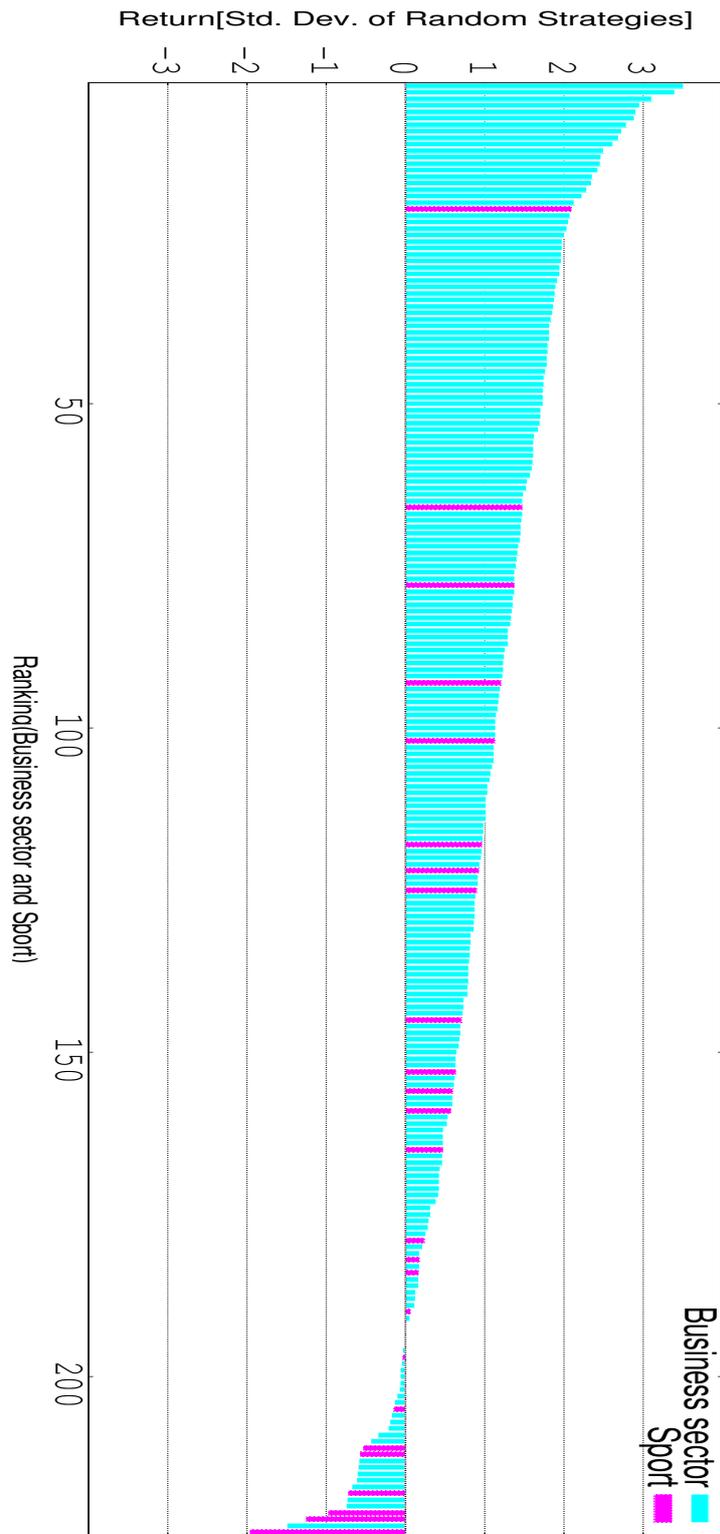


図6 Business sector と Sport & Sporting competition に分類されるトピックの累積収益率のランキング 青は Business sector に分類されるトピック，赤は Sport & Sporting competition に分類されるトピックの累積収益率を示す。

順位	トピック名	累積収益率
1	Metal / Mining	3.58
2	Entertainment Production	3.46
3	Food Processing	3.17
4	Food & Drug Retailing	3.01
5	Leisure & Recreation	2.97
6	Aluminum	2.94
7	Real Estate	2.84
8	Restaurants	2.78
9	Utilities	2.74
10	Media / Publishing	2.50

表2 表1の Business sector のニュースにおいて累積収益率が高いトピックトップ10

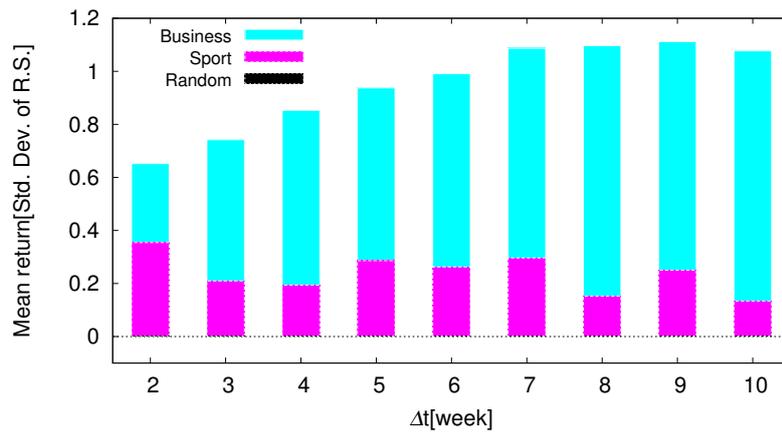


図7 Business sector と Sport & Sporting competition に属するトピックに関する過去  $\Delta t$  週間のニュースを利用した売買戦略による平均累積収益率  $\langle R \rangle$  青と赤は、それぞれ、Business sector と Sport & Sporting competition のニュースを利用した売買戦略による平均累積収益率  $\langle R \rangle$  を表す。ほとんど  $\langle R \rangle = 0$  を示す黒は、ランダムに売買したときの平均累積収益率である。

## 2.7 累積収益率の時期依存性

累積収益率の時期依存性を調べるために、2008年から2011年の各年の累積収益率の分布を観測した。図8は、Business sectorに属するトピックのニュースを利用してS&P500 indexを仮想的に売買したときの年間累積収益率の確率密度分布である。平均累積収益率 $\langle R \rangle$ は常に正の値を示す。また、大きな下落が起きている2009年と2011年は平均累積収益率が高い傾向にある。ランダムな売買による累積収益率の分布を帰無仮説とするKolmogorov-Smirnov検定では、帰無仮説は全ての年で有意水準1%で棄却される。つまり、時期ごとに累積収益率の差はあるが、ニュースを利用した売買戦略は常に有効である。

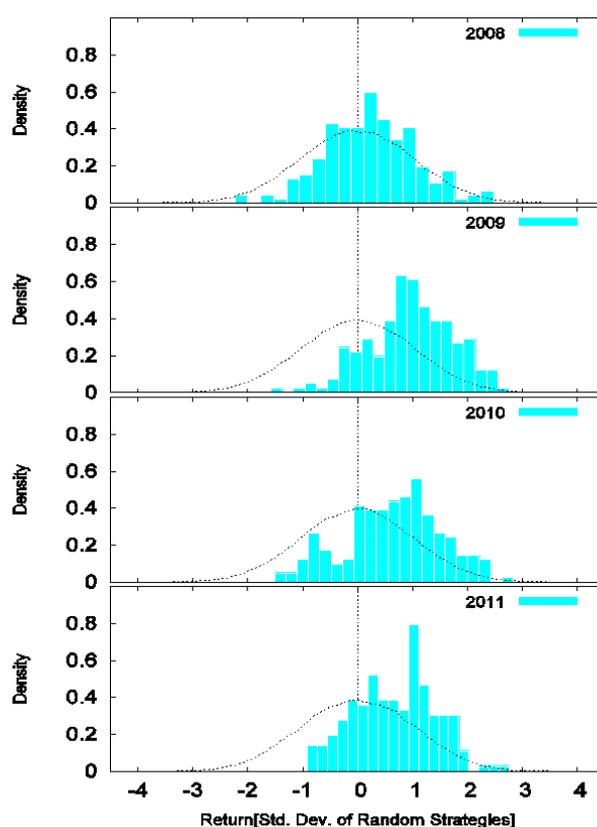


図8 Business sectorに属するトピックのニュースを利用してS&P500 indexを仮想的に売買したときの年間累積収益率 $R_i$ の確率密度分布 破線の分布は、ランダムな売買による累積収益率 $R_i$ の分布を示す。売買戦略のパラメータ $\Delta t$ に7週をセットした。年間の平均累積収益率 $\langle R \rangle$ は、2008年が0.22, 2009年が0.97, 2010年が0.66, 2011年が0.67である。Kolmogorov-Smirnov検定の $p$ 値は、2008年が $1.40 \times 10^{-3}$ , 2009年が $6.03 \times 10^{-35}$ , 2010年が $2.94 \times 10^{-15}$ , 2011年が $5.33 \times 10^{-16}$ である。

## 2.8 日本国債への応用

ニュース記事数の変動による市場価格変動の予測は株式市場以外でも可能であろう。図9は、2007年12月から2012年4月の期間で、Business sectorに関する日本語のニュース記事数を用いて日本国債10年物の金利を取引する売買戦略の累積収益率の確率密度分布である。<sup>\*5</sup> このニュースを利用した売買戦略の平均累積収益率は0.35である。ランダムな売買の累積収益率の分布とは有意に異なり、Kolmogorov-Smirnov 検定の  $p$  値は  $1.34 \times 10^{-5}$  である。ニュースを利用した売買戦略は日本国債に対しても有効であることが示された。

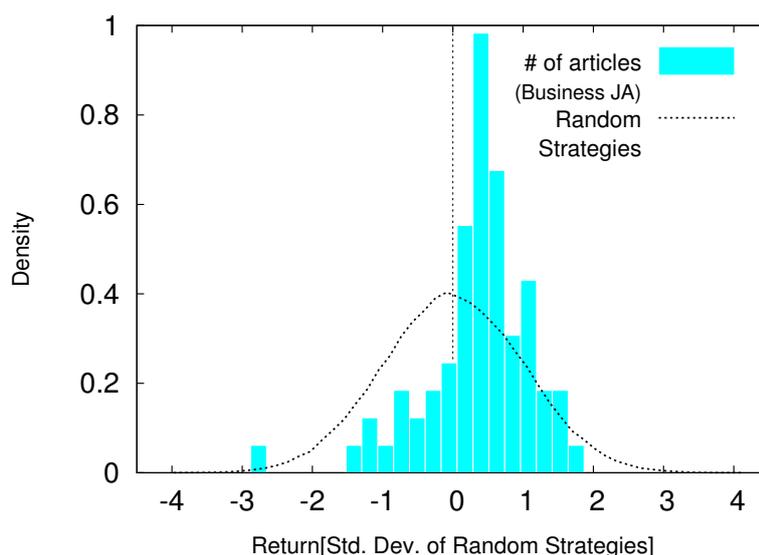


図9 Business sector の日本語ニュースを利用して日本国債10年物の金利を売買したときの累積収益率  $R$  の確率密度分布 破線の分布は、ランダムな売買による累積収益率  $R$  の分布を示す。売買戦略のパラメータ  $\Delta t$  に7週をセットした。平均累積収益率  $\langle R \rangle$  は、0.35である。破線の分布を帰無仮説としたKolmogorov-Smirnov 検定の  $p$  値は、それぞれ、 $1.34 \times 10^{-5}$  である。

## 2.9 まとめ

2章では、2007年12月から2012年4月までに THOMSON REUTERS 社が機関投資家向けに配信する英語のニュース915万記事と S&P500 index の日次時系列を用いて、ある株式市場に関連するニュース記事数の変動と将来のその市場の株式指数の変動に相関があることを示した。ある週に NYSE と NASDAQ に上場する銘柄に関連するニュース記事

<sup>\*5</sup> ただし、前節までと同様に、ニュース記事が130週より多く存在するトピックのみ分析に利用する。この条件に該当するトピック数は72個である。

数が増加/減少すると、その翌週に S&P500 index は下降/上昇する傾向にあるが、一方で、NYSE や NASDAQ と関連の低い他の市場の株式指数の変動はランダムに近い。また、企業の業績と密接に関係する Business sector のニュース記事数の変動には、市場は顕著に反応する。これらの傾向は各年で観測される。行動経済学のプロスペクト理論によると、投資家は損失回避を優先する。したがって、株価に下落のおそれがあるとき、投資家はその銘柄に関連する多くの情報を特に必要とし、報道機関はその要望に応じる。その結果として、ニュース量と株式指数の変動に関するこれらの統計的な特徴が現れると考えられる。

ニュースの記事数の増減データを利用した売買戦略は Web 上の検索数・閲覧数の増減を利用した売買戦略と同水準の収益が得られる。例えば、T. Preis らによれば、2008 年から 2011 年の Wikipedia の “Outline of economics” 中の “General economic concept” に分類される 285 ページの閲覧数を用いて Dow Jones Industrial Average を売買したときの平均累積収益率は、2008 年が 0.89、2009 年が 0.19、2010 年が 0.19、2011 年が 0.55 であった [15]。一方、本論文で示したように Business sector のニュースを利用した売買戦略の平均累積収益率は、2008 年が 0.22、2009 年が 0.97、2010 年が 0.66、2011 年が 0.67 であった。若干の傾向の違いは、投資家向けニュースから情報を得る機関投資家と Google や Wikipedia から情報を得る一般人での行動の違いを表していると考えられる。

### 3 日本国債・米国債・ドル円レートの3体相関

3章は以下のように構成される。3.1節では、国債先物と現物の比較を行う。3.2節では、日本国債先物の価格変動の統計性が株式市場や為替市場で観測される統計性と一致することを示す。3.3節では、日米間の国債の金利差とドル円レートに強い相関があることを示す。ただし、米国債の金利とドル円レートの相関に対し、日本国債の金利とドル円レートの相関は弱い。3.4節では、日本銀行による国債買入が行われた日に注目すると、日本国債とドル円レートに強い相関が現れることを示す。

国債の金利と価格の関係は以下の式で決まっており、価格が上昇/下降するとき金利は下降/上昇する。国債先物の場合は、利率と残存年数が固定されており、価格と金利が完全に連動するので、価格の変動をみることと金利の変動をみことは同じことである。

$$\text{価格} = \frac{100 + \text{利率} \times \text{残存年数}}{100 + \text{金利} \times \text{残存年数}} \times 100$$

#### 3.1 日本国債の先物と現物の連動性

国債先物は長期国債先物取引と呼ばれるもので、残存期間が10年の標準物と呼ばれる架空の国債に対して先物取引を行う。先物は頻繁に取引が行われていて流動性が高いため、市場の動向をつかむ上で重要な材料と言われている。しかし、その一方で、現物の動きを反映していないのではないかという指摘もある。まず3.1節で、日本国債の先物と現物が連動していることを示し、以降の節では流動性の高い先物に対して分析を行う。

図10は、2012年12月から2013年12月までの先物と現物の価格の分次時系列である。黒色は先物、赤色と緑色と青色と桃色は、それぞれ現物の10年325回債と10年327回債と10年328回債と10年329回債の価格を示す。先物と現物が連動した値動きをしていることがわかる。図11は、先物と現物の10年325回債の正の価格変動及び負の価格変動の絶対値

$$|\Delta p(t)| = |p(t + \tau) - p(t)| \quad (15)$$

の5分刻み( $\tau = 5\text{days}$ )の累積分布である。赤色と青色はそれぞれ正と負の価格変動、太線と細線はそれぞれ先物と現物を示す。点線は標準正規分布を示す。正と負の価格変動、どちらの場合でも、先物と現物の分布はほとんど同じ分布になっており、また正規分布に近い分布になっていることがわかる。これらの結果は、先物と現物の価格変動が連動していることを示唆する。

図12は、先物と現物の取引の約定数の日次時系列である。先物と現物の2つの時系列の相関係数は高く、0.58である。約定数についても、先物と現物は連動している。

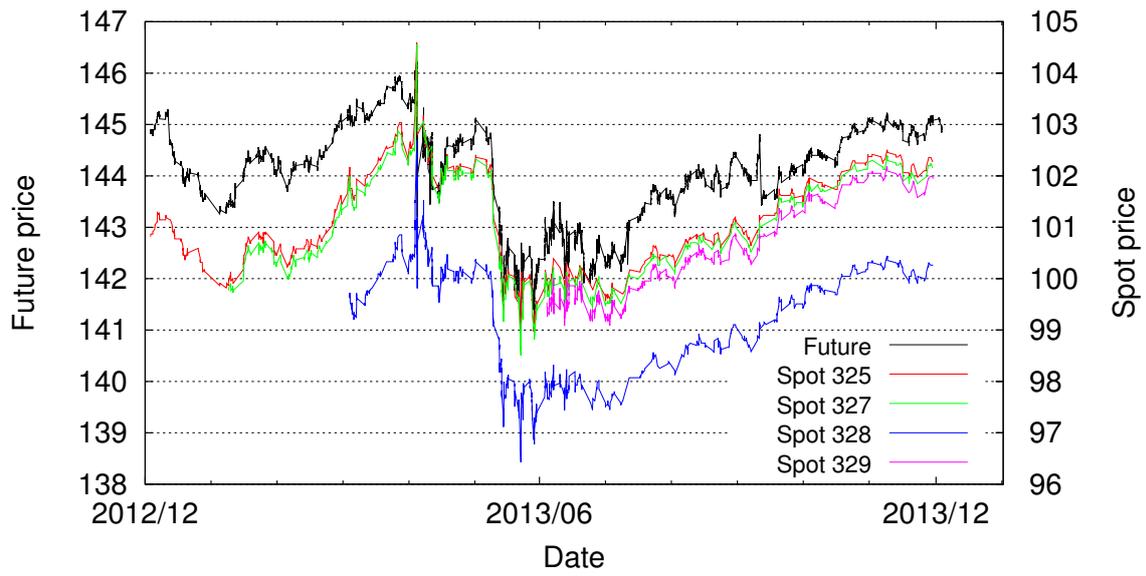


図 10 日本国債先物と現物の価格の分次時系列 黒色は先物，赤色と緑色と青色と桃色は，それぞれ現物の 10 年 325 回債と 10 年 327 回債と 10 年 328 回債と 10 年 329 回債の価格を示す。

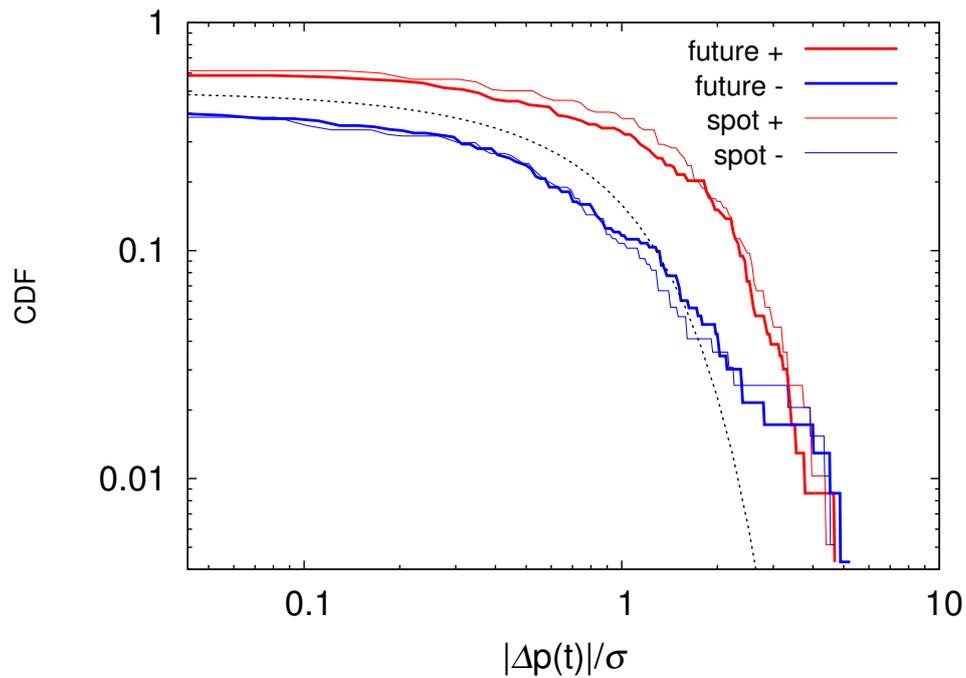


図 11 日本国債先物と現物の 5 分刻みの価格変動の累積分布 赤色と青色はそれぞれ正と負の価格変動を，太線と細線はそれぞれ先物と現物を示す。点線は標準正規分布を示す。

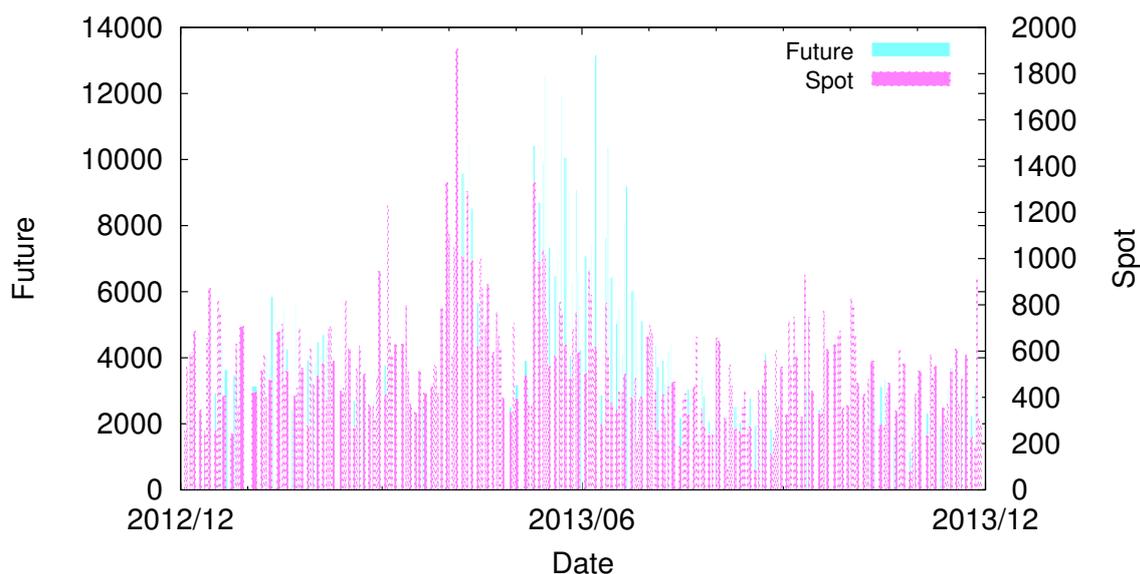


図 12 日本国債先物と現物の約定数の日次時系列 青色と赤色は、それぞれ先物と現物を示す

### 3.2 日本国債先物の統計性

株式市場や為替市場でみられる統計性として、価格変動の累積分布が短い時間スケールではべき分布に近い分布になり、長い時間スケールでは正規分布に近い分布になることや、価格変動の標準偏差が時間スケールの 0.5 乗に比例することが知られている。これらの統計性が日本国債先物についても同様にみられるか確認する。

図 13 は日本国債先物の正の価格変動及び負の価格変動の絶対値の、10000 分刻み ( $\tau = 10000\text{min}$ ) の場合と 10 分刻み ( $\tau = 10\text{min}$ ) の場合の累積分布である。10000 分は約 1 週間である。図の左側の破線は標準正規分布を、図の右側の破線はべき指数  $-3$  のべき分布を示す。 $\tau = 10\text{min}$  の場合は価格変動の累積分布がべき分布に近い分布になっており、 $\tau = 10000\text{min}$  の場合は価格変動の累積分布が正規分布に近い分布になっている。日本国債においても、価格変動の累積分布が、短い時間スケールでべき分布に、長い時間スケールで正規分布に近い分布となる。

図 14 は日本国債先物の価格変動の標準偏差

$$\sigma(\tau) = \sqrt{\langle (p(t+\tau) - p(t))^2 \rangle - \langle p(t+\tau) - p(t) \rangle^2} \quad (16)$$

と時間スケール  $\tau(\text{min})$  の関係を示している。この図の破線はべき指数 0.5 を示しており、日本国債先物においても価格変動の標準偏差が時間スケールの 0.5 乗に比例していることがわかる。この結果は、日本国債の価格がランダムウォーク過程に従うことを意味する。

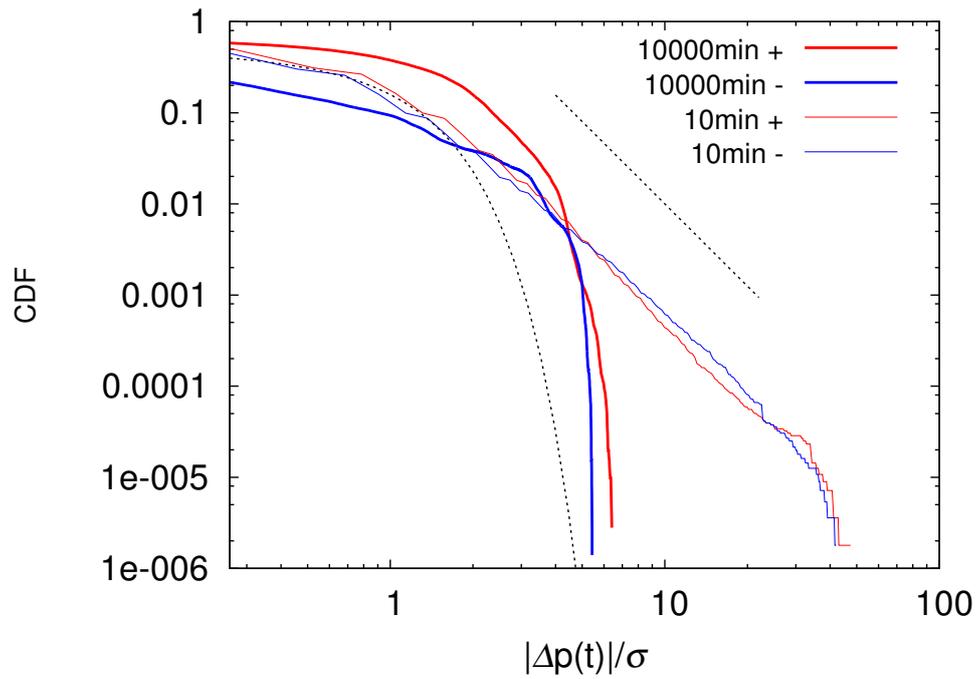


図 13 日本国債先物の価格変動の累積分布 赤色と青色はそれぞれ正と負の価格変動を，太線と細線はそれぞれ  $\tau = 10000\text{min}$  と  $\tau = 10\text{min}$  を示す．左側の破線は標準正規分布，右側の破線は指数  $-3$  のべき分布を示す．

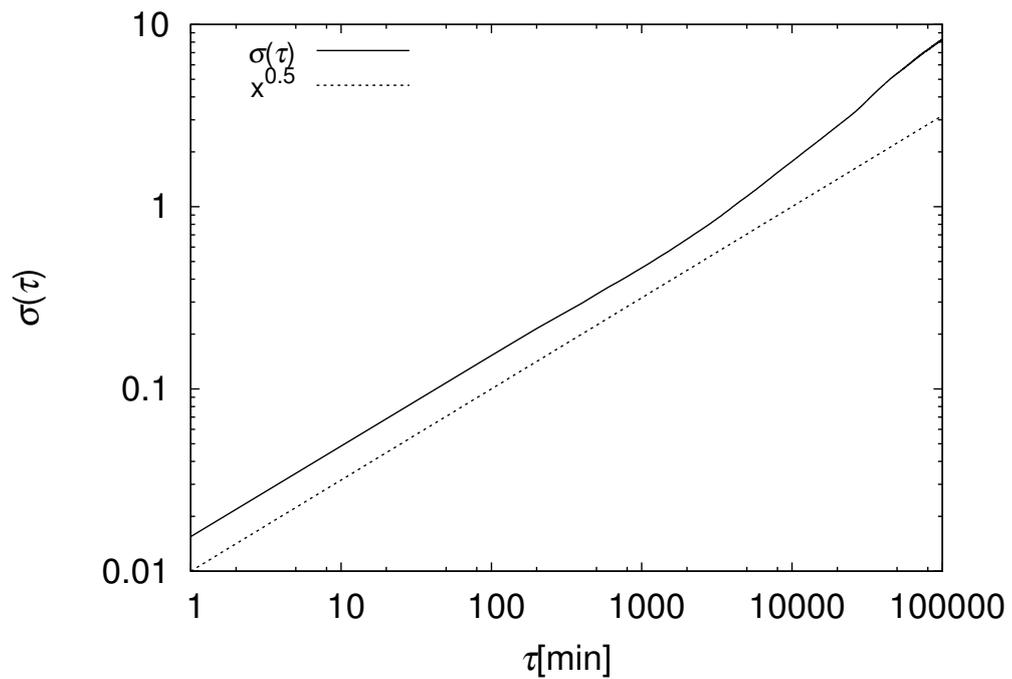


図 14 時間スケール  $\tau$  に対する価格変動の標準偏差 破線はべき指数  $0.5$  を示している．

### 3.3 日本国債・米国債とドル円レートの関係

日本国債と米国債の金利差とドル円レートに相関があることを示す。また、日本国債と米国債、どちらの方がドル円レートに影響を与えているか明らかにする。図 15 は日本国債  $i(t)$  と米国債の金利  $i^*(t)$  とドル円レート  $S(t)$  の日次時系列である。金利平価説より国債金利が高い方の国の通貨が将来的に減価するので、日米間の国債の金利差  $i(t) - i^*(t)$  とドル円レートの対数変動  $\ln \frac{S(t+\tau)}{S(t)}$  には強い相関が生じると考えられる\*6。図 16 は  $\tau = 586$  日の場合の日米間の国債の金利差とドル円レートの対数変動の時系列である。たしかに、2つの時系列の相関係数は高く、0.56 である。

$i(t) - i^*(t)$  の 1 日あたりの変動

$$(i(t+1) - i^*(t+1)) - (i(t) - i^*(t)) \quad (17)$$

と  $\ln \frac{S(t+\tau)}{S(t)}$  の 1 日あたりの変動

$$\ln \frac{S(t+\tau+1)}{S(t+1)} - \ln \frac{S(t+\tau)}{S(t)} \quad (18)$$

の  $\tau$  ごとの相関係数を測る。ここで、

$$\begin{aligned} \Delta i(t) &= i(t+1) - i(t) \\ \Delta i^*(t) &= i^*(t+1) - i^*(t) \\ \Delta \ln \frac{S(t+\tau)}{S(t)} &= \ln \frac{S(t+\tau+1)}{S(t+1)} - \ln \frac{S(t+\tau)}{S(t)} \end{aligned} \quad (19)$$

とおく。図 17 の黒線は  $\Delta i(t) - \Delta i^*(t)$  と  $\Delta \ln \frac{S(t+\tau)}{S(t)}$  の相関係数を示している。日米間の国債の金利差とドル円レートの対数変動の相関は高い。赤線と青線はそれぞれ、米国債の金利  $\Delta i^*(t)$  を定数  $\langle \Delta i^*(t) \rangle$  とした場合、日本国債の金利  $\Delta i(t)$  を定数  $\langle \Delta i(t) \rangle$  とした場合の相関係数を示している。米国債の金利を定数とした場合、すなわち日本国債の金利とドル円レートの対数変動のみに注目した場合は、相関が弱いことがわかる。一方、米国債の金利とドル円レートの対数変動のみに注目した場合は、相関が強い。これらの結果は、日米間の国債の金利差とドル円レートの対数変動は強い相関関係にあるが、日本国債の金利がドル円レートに与える影響は弱く、米国債の金利による影響が強いことを示唆する。図 15 より、日本国債の金利はかなり低く、これ以上ほとんど金利が下がらないことがわかる。これは、米国債に比べて日本国債がドル円レートに与える影響が小さい原因の 1 つと考えられる。

\*6 ただし、国債金利は年利から  $\tau$  日の金利になおす。

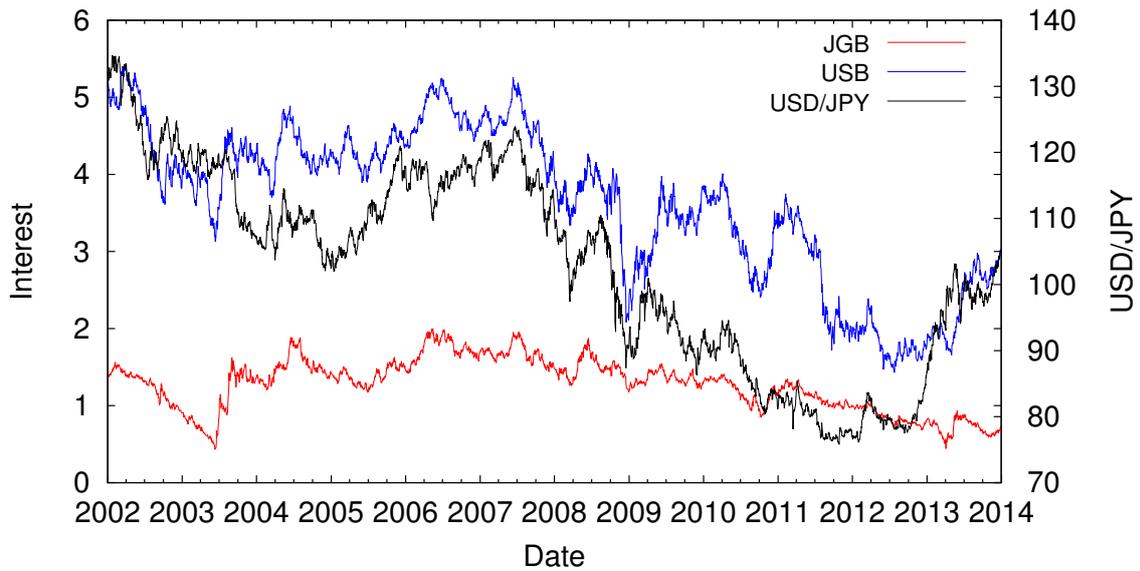


図 15 日本国債・米国債の金利とドル円レートの日次時系列 赤線は日本国債の金利，青線は米国債の金利，黒線はドル円レートの時系列を示す。

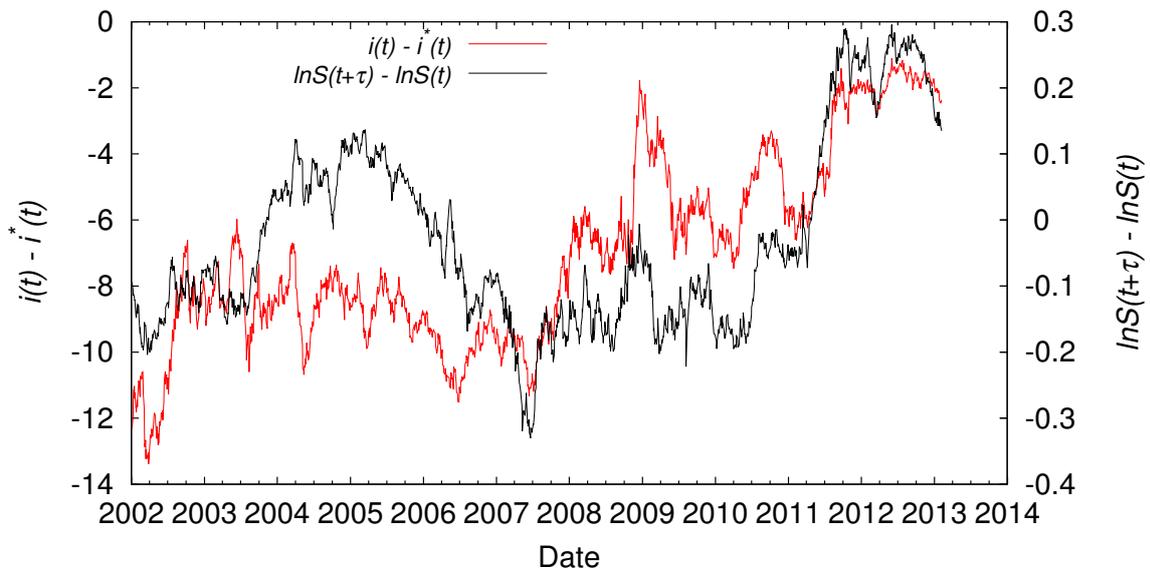


図 16  $\tau = 586$  日の場合の日米間の国債の金利差とドル円レートの対数変動の日次時系列 赤線が日米間の国債の金利差，黒線がドル円レートの対数変動を示している。

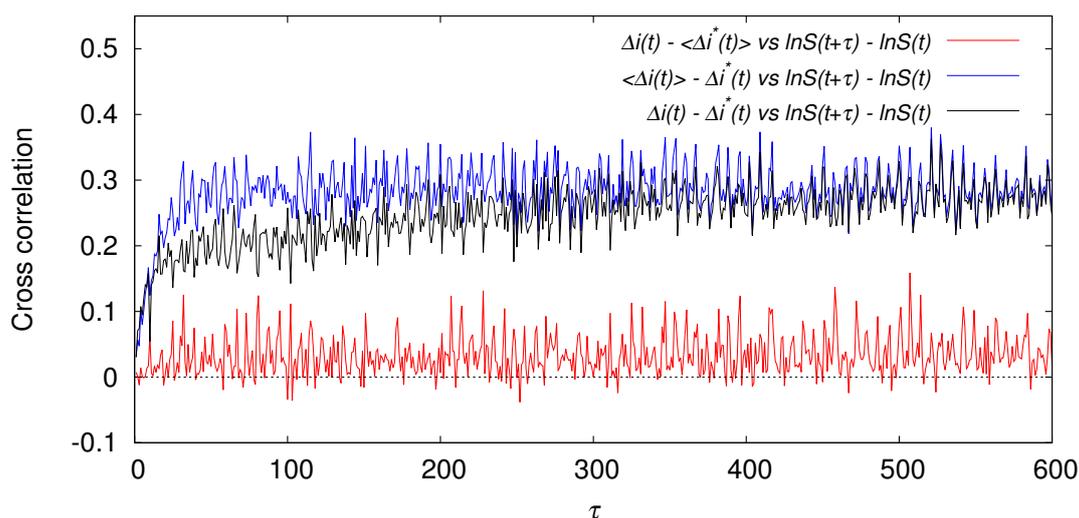


図 17 日米間の国債の金利差とドル円レートの対数変動の相関係数 黒線は日本国債・米国債とドル円レートの相関係数，赤線は日本国債とドル円レートの相関係数，青線は米国債とドル円レートの相関係数を示す。

### 3.4 日本国債とドル円レートに対する日本銀行の国債買入の影響

3.3 節では，日米間の国債の金利差とドル円レートの対数変動は強い相関関係にあるが，米国債に比べて，日本国債の金利がドル円レートに与える影響が弱いことが示された。では，常に日本国債の影響がほとんどなく，米国債のみがドル円レートを動かすのか。3.4 節では，日本銀行による国債買入が行われた日に，日本国債に価格上昇（金利下降）のトレンド，ドル円レートに円高方向のトレンドが発生し，また，日本国債とドル円レートの相関が強くなることを示す。

日本銀行の国債買入は次のような手順で行われる。まず，日本銀行が国債買入を行うというアナウンスを行い，同時にその金額を提示する。このアナウンスが行われた日をオファード，提示された金額をオファード額と呼ぶ。それに対し，複数の銀行や保険会社が応札する。複数の銀行や保険会社によって応札された金額の合計を応札額と呼ぶ。オファード日から 3 営業日後に日本銀行によって，どの銀行・保険会社からそれぞれいくら買入れるか決められ，実際に取引が行われる。この日を落札日，買入の合計金額を落札額と呼ぶ。落札額はオファード額とほぼ同額で，おおよそ 100 億円から 1000 億円単位となっている。この一連の流れを国債買入オペと呼ぶ。図 18 はオファード額の月次時系列である。2012 年にアベノミクスによる金融緩和が行われてから，オファード額が急激に増えていることがわかる。

オファード日以降に，日本国債の価格とドル円レートに変動があるか明らかにする。図 19 の赤線はオファード日の前日の終値を基準にした前後 3 日間の価格変動の平均値を示す。

$\tau = 1$  日に、日本国債の価格は上昇し、ドル円レートは円高方向に変動していることがわかる。つまり、オファー日に日本国債とドル円レートが変動したということになる。日本銀行が日本国債を買い入れるというアナウンスをすることで、市場における日本国債の売り注文が減り、日本国債の価格が上昇したと考えられる。また、日本国債の価格上昇(金利下降)により、ドル円レートが円高方向へ変動したと考えられる。

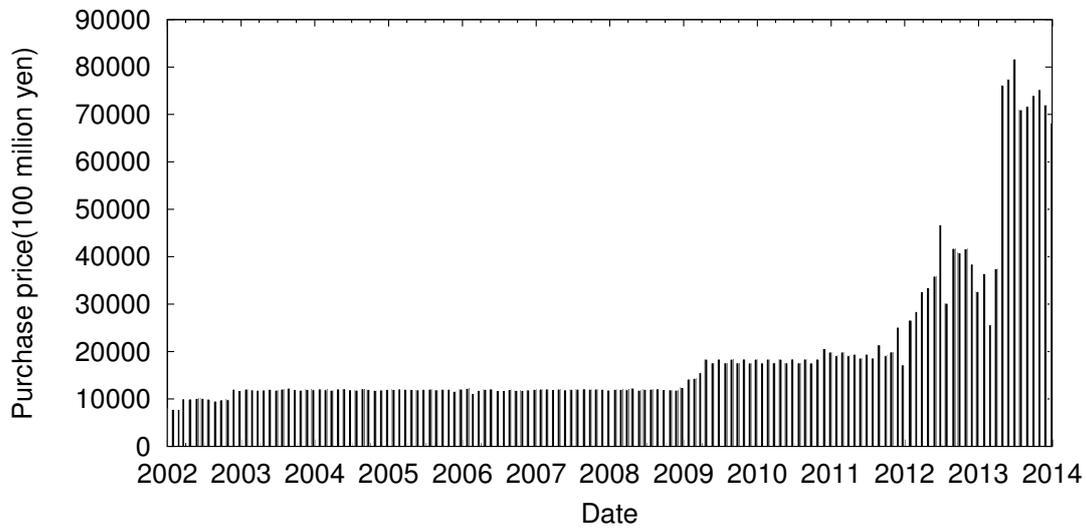


図 18 日本銀行の国債買入のオファー額の月次時系列 縦軸は 1 億円単位.

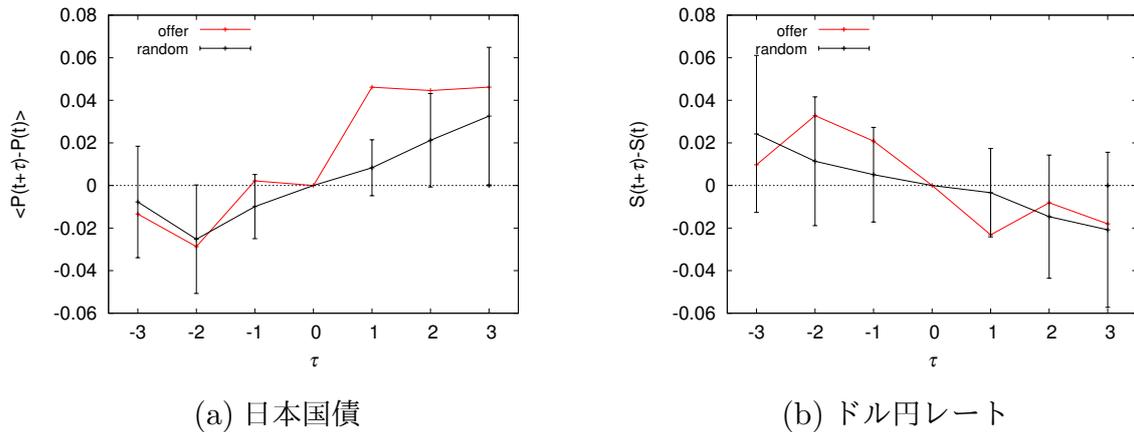


図 19 オファー日の前後 3 日間の価格変動 赤線はオファー日の前日の終値を基準にした前後 3 日間の価格変動の平均値を示す。エラーバーは、国債買入があった日数と同じ日数だけランダムに日付を選択し、その前後 3 日間の価格変動の平均値を計算する操作を 1000 回行い、その平均値と標準偏差を計算することでプロットした。

次に、オファー日前後だけではなく、オファー日の日中に日本国債とドル円レートにトレンドが生じていることを示す。図 20(a) はオファー日の、図 20(b) はオファーがなかった日

の、日中の日本国債とドル円レートの分次時系列を 100 を初期値として繋ぎ合わせてプロットしたものである。赤線は日本国債、黒線はドル円レートを示す。図 20(a) より、オファー日に、日本国債の価格が上昇、ドル円レートが円高方向へ変動していることが読み取れる。一方、オファーがなかった日に、そのような特徴は見られない。図 21 の青線と赤線は、それぞれ、図 20 の (a) と (b) における日本国債の価格変動  $P(t+\tau) - P(t)$  とドル円レートの変動  $S(t+\tau) - S(t)$  の相関係数を示している。オファーがなかった日に比べて、オファー日の日本国債の価格変動とドル円レートの変動の相関は強い。

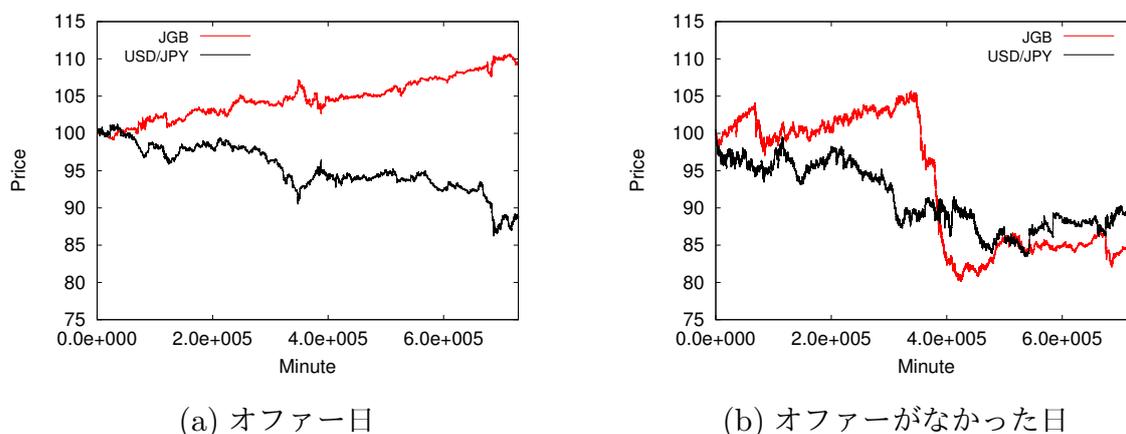


図 20 日本国債とドル円レートの日中の時系列 赤線は日本国債の価格、黒線はドル円レートを示す。

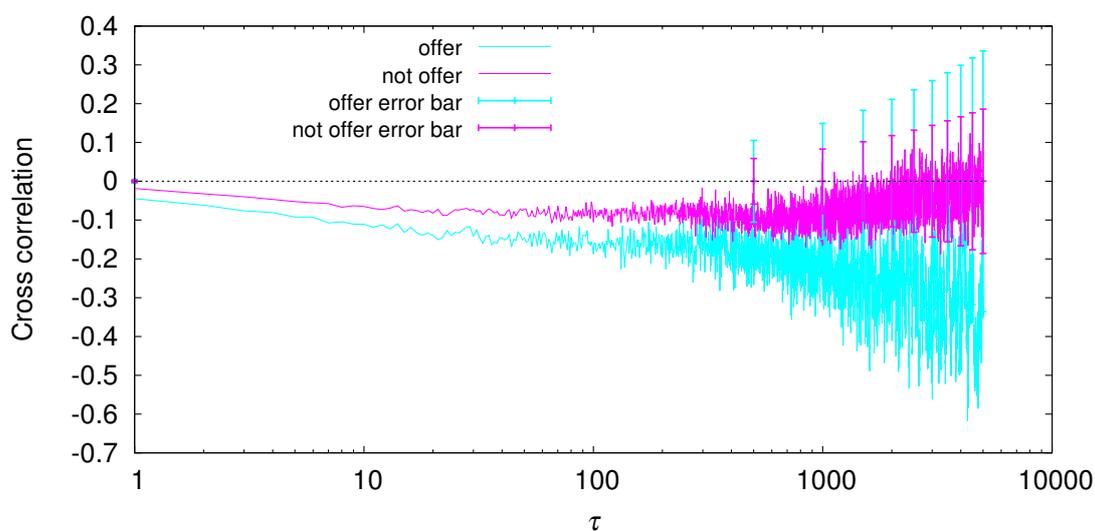


図 21 日本国債の価格変動とドル円レートの変動の相関係数 青線はオファー日、赤線はオファーがなかった日の、日本国債の価格変動とドル円レートの変動の相関係数を示す。相関係数の誤差の値は  $0 \pm 1.96/n^{0.5}$  で定義される ( $n$  はサンプル数を表す)。

図 22(a) はオファー日の、図 22(b) はオファーがなかった日の日本国債の価格変動  $P(t + \tau) - P(t)$  の累積分布である。  $\tau = 10$  分の時間スケールの場合、オファーがあった日となかった日で分布の違いはほとんどない。しかし、  $\tau = 1500$  分 (約 1 日) の時間スケールの場合、オファーがなかった日は正の価格変動に比べて負の価格変動の分布の裾野が広がっているのに対して、オファーがあった日は正の価格変動の分布の方が裾野が広がっている。この結果は、オファーがあった日に日本国債に価格上昇のトレンドが発生していることを示唆する。

図 23(a) はオファー日の、図 23(b) はオファーがなかった日のドル円レートの変動  $S(t + \tau) - S(t)$  の累積分布である。日本国債と同様に、  $\tau = 1500$  分 (約 1 日) の時間スケールの場合に、オファーがあった日となかった日の分布の違いが見られる。オファーがあった日は、なかった日と比べて、円高方向の変動の分布の裾野が広がっている。この結果は、オファーがあった日にドル円レートに円高方向のトレンドが発生していることを示唆する。

図 24(a) は、国債買入額 (オファー額) と日本国債の  $\tau = 5$  日の価格変動の散布図である。相関係数は低く、  $-6.3 \times 10^{-3}$  である。図 24(b) はドル円レートの場合で、日本国債の場合と同様に相関係数は低く、  $-6.6 \times 10^{-2}$  である。国債買入のオファーがあった日に、日本国債とドル円レートの両方にトレンドが発生するが、国債買入額 (オファー額) の影響はない。つまり、オファーがあったか否かが、トレンドの発生を決めるということである。

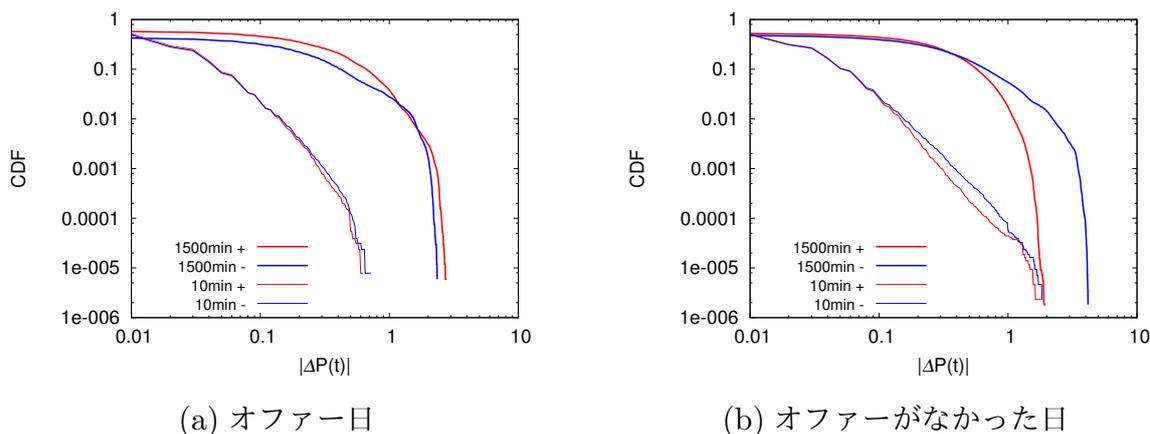
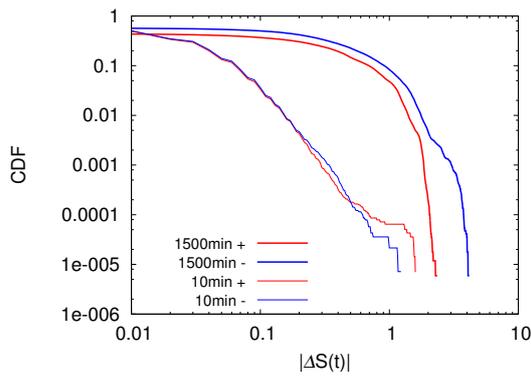
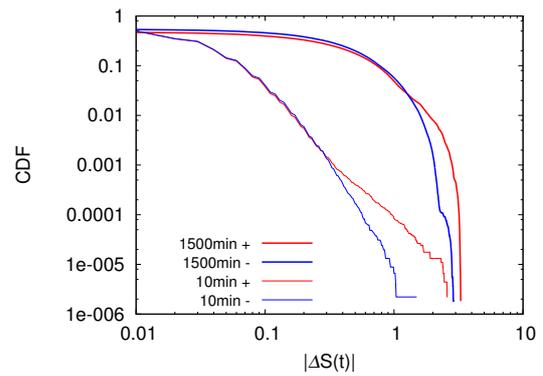


図 22 日本国債の価格変動の累積分布 赤色と青色はそれぞれ正と負の価格変動を、太線と細線はそれぞれ  $\tau = 10000\text{min}$  と  $\tau = 10\text{min}$  を示す。

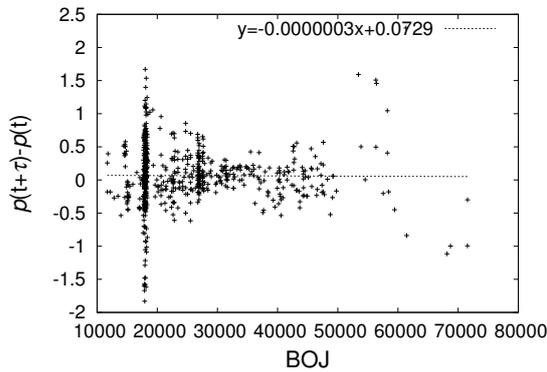


(a) オファー日

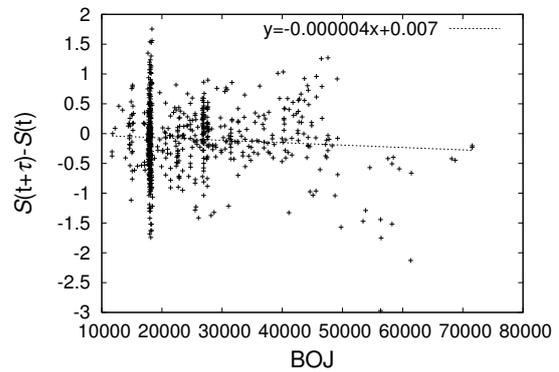


(b) オファーがなかった日

図 23 ドル円レートの変動の累積分布 赤色と青色はそれぞれ正と負の価格変動を、太線と細線はそれぞれ  $\tau = 10000\text{min}$  と  $\tau = 10\text{min}$  を示す。



(a) 日本国債



(b) ドル円レート

図 24 オファー額と日本国債・ドル円レート 点線は近似線を示す。

### 3.5 まとめ

3章では、日本銀行の国債買入のオファー日に、日本国債に価格上昇(金利下降)のトレンドが発生し、ドル円レートとの相関が強くなることを示した。経済学における金利平価説によると、2国間の金利差により為替レートが決定する。たしかに、日米間の国債の金利差とドル円レートの対数変動に強い相関はあるが、日本国債による影響は少ない。これは、日本国債の金利が低く、これ以上ほとんど下がらないことが原因と考えられる。しかし、日本銀行の国債買入のオファー日に注目すると、日本国債の価格は上昇(金利下降)し、また、ドル円レートも円高方向に変動し、日本国債とドル円レートの相関が強くなる。この結果は、日本銀行の国債買入オペレーションには、日本国債の価格を上昇させ、それによりドル円レートも変動させる効果があることを示唆する。

## 付録 A 株式銘柄に関するニュースの分類

本論文で使用した THOMSON REUTERS のニュース記事には株式銘柄やトピックのコードが付随しており、それを元にニュース記事の分類を行った。このようなコードがない他の報道機関のニュースについても同様の分類ができるように、株式銘柄コードごとにニュース本文の単語の重要度について調査する。重要度の高い単語が本文中に存在するニュースをその株式銘柄に関するニュースとする。1つの例として、2007年から2008年の American International Group (AIG) の株式銘柄コードが付いた英語ニュース記事の ALERT ニュースと HEADLINE ニュースのタイトル部分を対象に分析する。

単語の重要度として、tf 値 (単語の出現頻度) と idf 値 (逆文書頻度) を計算する。tf 値と idf 値は以下のように計算できる。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (20)$$

$$idf_i = \ln \frac{|D|}{|\{d : d \ni t_i\}|} \quad (21)$$

ここで、 $n_{i,j}$  は単語  $i$  の株式銘柄コード  $j$  の文書における出現回数、 $|D|$  は総文書数、 $|\{d : d \ni t_i\}|$  は単語  $i$  を含む文書数である。ある株式銘柄コードのニュース記事の本文中に頻繁に出現する単語は tf 値が高くなる。しかし、一般的な単語 (例: on, to) のように多くの文書に出現する単語は idf 値が低くなる。idf は特定の文書にしか出現しない単語の重要度を上げる。

ニュース本文を単語ごとに分割する際に、1単語ごとだけではなく、2単語ごと、3単語ごとでも分割を行う。例えば、“NYSE ORDER IMBALANCE <AIG.N> 209000 SHARES ON BUY SIDE” という文章を2単語ごとに分割する場合は“NYSE ORDER”, “ORDER IMBALANCE”, “IMBALANCE <AIG.N>”, “<AIG.N> 209000”, “209000 SHARES”, “SHARES ON”, “ON BUY”, “BUY SIDE” と分割する。3単語ごとの分割も同様に行う。2単語ごと、3単語ごとでも分割するのは、株式銘柄によって隣り合う頻度が高い単語の組み合わせが存在するからである。例えば、AIG の正式名称 “American International Group” の中に含まれる単語は頻度が高くなるはずなので tf 値は大きくなる。しかし、“International” や “Group” は他の銘柄のニュースでも多く出てくるため idf 値が低くなってしまう。ここで、2単語ごと、3単語ごとでも分割を行うと “American International” や “American International Group” といった単語の組み合わせは他の銘柄にはほとんどないので idf 値も大きくなる。

図 25 は AIG のコードがついたニュース本文に出現する1単語の tf 値に対する idf 値の散布図である。tf 値は対数をとっている。破線は、対数をとった tf 値で15分割し、idf 値

の平均値を計算することでプロットした。tf 値が低い/高い単語ほど idf 値は高い/低い傾向にあることがわかる。この平均値より  $h$  倍以上高い単語が出現するニュース記事を AIG に関するニュースとする。

ニュース記事の分類の精度を測るために、“Type 1 error”と“Type 2 error”を調べる。“Type 1 error”は実際は AIG に関するニュースではないのに、分類されてしまうエラーである。AIG のコードがついていないニュースをランダムに 300 個とりだす。その 300 個のニュースの中に AIG に関するニュースがないか人為的に調べ、もしあった場合は取り除き、AIG と関係のないニュースのみにする。指定した単語が入ったニュースの割合が“Type 1 error”である。“Type 2 error”は実際は AIG に関するニュースなのに、AIG に関するニュースとして分類されないエラーである。AIG ニュースのコードがついたニュースの中で、指定した単語が入っていないニュースの割合がエラーとなる。この 2 つのエラーが小さくなるようにしきい値  $h$  を決める。このしきい値  $h$  は 1 単語ごとの分割、2 単語ごとの分割、3 単語ごとの分割、それぞれの場合で決める。

1 単語ごとの場合のしきい値  $h = 1.5$ 、2 単語ごとの場合のしきい値  $h = 1.25$ 、3 単語ごとの場合のしきい値  $h = 1.1$  としたとき、“Type 1 error”と“Type 2 error”は共に小さくなり、それぞれ 0%、31.47% となる。このときの 1 単語の数は 35 個、2 単語の数は 403 個、3 単語の数は 415 個である。これらの単語が本文中に存在するニュース記事を AIG のニュースとして分類できる。

他の株式銘柄コード、またはトピックコードについても同様の分析をすることで、その株式銘柄・トピックに関するニュースとして分類することができる。

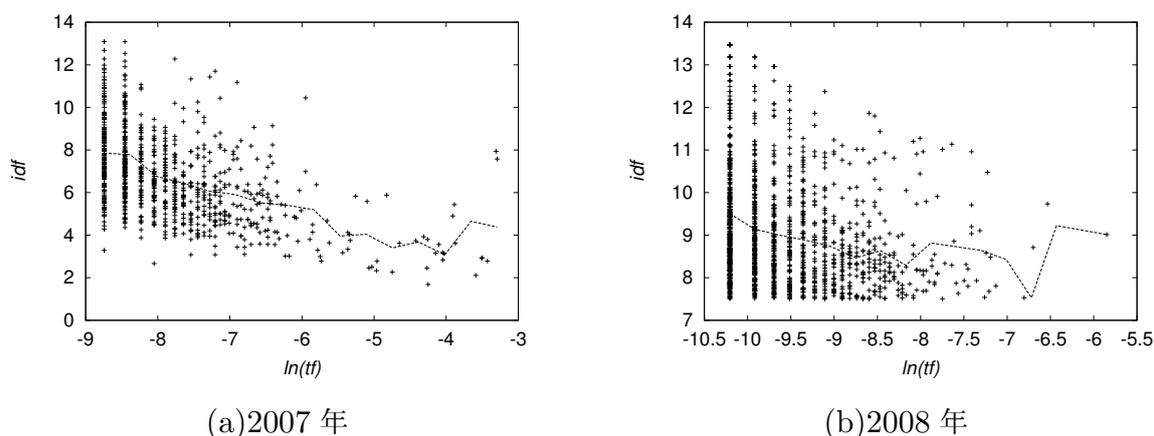


図 25 AIG のコードがついたニュース記事の本文中に出現する単語の、tf 値に対する idf 値の散布図 (a) は 2007 年のニュース記事を使用した場合。(b) は 2008 年のニュース記事を使用した場合。

## 謝辞

本論文を執筆するにあたり，多くの方々にお世話になりました．ここに感謝の意を述べたいと思います．

国立情報学研究所の水野貴之先生には丁寧なご指導を頂き，また研究会や国際会議での発表の機会を多く与えて下さり，大変感謝しております．指導教授である香取眞理先生には，学部生時代からご指導を頂き，研究に対する姿勢など多くのことを学ばせて頂きました．深く感謝致します．東京大学大学院情報理工学系研究科の大西立顕先生と東京大学大学院経済学研究科の渡辺努先生には，本論文の後半部分にあたる日本国債に関する研究で多くの助言を頂き，大変感謝しております．香取研究室の高橋幹君と本田良二郎君との議論の中から，研究に関する多くのヒントを得ることができました．感謝致します．杉本秀彦先生と若林淳一先生には本論文の副査をして頂きました．深く感謝致します．

## 参考文献

- [1] Z. Q. Jiang, L. Guo & W. X. Zhou, Endogenous and exogenous dynamics in the fluctuations of capital fluxes, *European Physical Journal B* **57**(3), 347-355 (2007).
- [2] D. Sornette, Endogenous versus Exogenous Origins of Crises, in *Extreme events in nature and society*, ed. S. Albeverio, V. Jentsch & H. Kantz, (pp. 95-119), Springer, Berlin Heidelberg (2006).
- [3] M. E. Newman, Power laws, Pareto distributions and Zipf's law, *Contemporary Physics* **46**(5), 323-351 (2005).
- [4] T. Mizuno, T. Ohnishi & T. Watanabe, Novel and topical business news and their impact on stock market activities, arXiv preprint arXiv:1507.06477 (2015).
- [5] R. Hisano, D. Sornette, T. Mizuno, T. Ohnishi & T. Watanabe, High quality topic extraction from business news explains abnormal financial market volatility, *PLoS ONE* **8**(6), e64846 (2013).
- [6] M. Alanyali, H. S. Moat & T. Preis, Quantifying the relationship between financial news and the stock market, *Scientific Reports* **3**, 3578 (2013).
- [7] T. Mizuno, K. Takei, T. Ohnishi & T. Watanabe, Temporal and cross correlations in business news, *Progress of Theoretical Physics Supplement* **194**, 181-192 (2012).
- [8] A. M. Petersen, F. Wang, S. Havlin & H. E. Stanley, Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws, *Physical Review E* **82**(3), 036114 (2010).
- [9] J. G. Rangel, Macroeconomic news, announcements, and stock market jump intensity dynamics, *Journal of Banking & Finance* **35**(5), 1263-1276 (2011).
- [10] J. Bollen, H. Mao & X. Zeng, Twitter mood predicts the stock market, *Journal of Computational Science* **2**(1), 1-8 (2011).
- [11] R. P. Schumaker & H. Chen, Textual analysis of stock market prediction using breaking financial news: The AZFin text system, *ACM Transactions on Information Systems (TOIS)* **27**(2), 12 (2009).
- [12] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai & A. Kappas, Sentiment strength detection in short informal text, *Journal of the American Society for Information Science and Technology* **61**(12), 2544-2558 (2010).
- [13] A. Tversky & D. Kahneman, Loss aversion in riskless choice: A reference-dependent model, *Quarterly Journal of Economics* **106**(4), 1039-1061 (1991).

- [14] T. Preis, H. S. Moat & H. E. Stanley, Quantifying trading behavior in financial markets using Google Trends, *Scientific Reports* **3**, 1684 (2013).
- [15] H. S. Moat, C. Curme, A. Avakian, D. Y. Kenett, H. E. Stanley & T. Preis, Quantifying Wikipedia usage patterns before stock market moves, *Scientific Reports* **3**, 1801 (2013).
- [16] C. Curme, T. Preis, H. E. Stanley & H. S. Moat, Quantifying the semantics of search behavior before stock market moves, *Proceedings of the National Academy of Sciences* **111**(32), 11600-11605 (2014).
- [17] Y. Tahira & T. Mizuno, Amount of news before stock market fluctuations, 投稿準備中.
- [18] T. Mizuno, Y. U. Saito, T. Watanabe & H. Takayasu, Characteristic market behaviors caused by intervention in a foreign exchange market, in *Practical Fruits of Econophysics*, ed. H. Takayasu, (pp. 33-37), Springer, Tokyo (2006).
- [19] Thomson Reuters, Reuters Newscope Sentiment Engine:Guide to sample data and system overview, *Thomson Reuters White Paper* (2008).
- [20] Thomson Reuters, Reuters Newscope Sentiment Engine:Output image and file format, *Thomson Reuters White Paper* (2008).
- [21] 沢木勝茂, ファイナンスの数理, 朝倉書店 シリーズ [現代人の数理] 8 (1994).